DIALOGO
Multidisciplinary JOURNAL
Available online at www.dialogo-conf.com/dialogo-journal/

This paper was presented at
The Annual Virtual International Conference on the Dialogue between Science and Theology (DIALOGO-CONF 2020 VIC)
held online,
from November 3 - 12, 2020

Prediction of Big Data Analytics (BDA) on Social Media: Empirical Study

Ahed J Alkhatib
1 Department of Legal Medicine, Toxicology and Forensic Medicine, Jordan University of Science & Technology,
2 International Mariinskaya Academy, department of medicine and critical care, department of philosophy, Academician secretary of the Department of Sociology.
JORDAN
ajalkhatib@just.edu.jo

Shadi Mohammad Alkhatib
Department of Software Engineering,
Hashemite University,
Zarqa, JORDAN
shadimalkhatib@gmail.com

Hani Bani Salameh
Department of Software Engineering,
Hashemite University,
Zarqa, JORDAN
hani@hu.edu.jo

ARTICLE INFO

Article history:
Received 15 October 2020
Received in revised form 30 October
Accepted 31 October 2020
Available online 30 November 2020
doi: 10.18638/dialogo.2020.7.1.19

Keywords:
Big Data; social media; neural network; data clustering;

ABSTRACT

Currently, most studies are moving towards Big Data Analytics (BDA) because they are important in research, and this is becoming increasingly important as Internet and Web 2.0 technologies become increasingly popular and how to handle this massive data. Moreover, this proliferation of the Internet and social media has revolutionized the search process. With this Big Data of data generated by users using social media or electronic platforms, the use of these details and daily activities is integrated with tools designed for analysis. The topic of analyzing big social media will be discussed and an intensive explanation will be given to the topic of Big Data. This paper compares Big Data analysis techniques using several methods of analysis, the first technique using neural networks and the second technique using data clustering. The purpose of this study is to infer the ages that use social media and what are their interests in writing and in the end, who are the most widely used social media males or females.

Citation: Alkhatib, Ahed J, Shadi Mohammad Alkhatib, and Hani Bani Salameh. "Prediction of Big Data Analytics (BDA) on Social Media: Empirical Study." *DIALOGO*, ISSN: 2393-1744, vol. 7, issue 1 (November 2020): 23-31. DOI: 10.18638/dialogo.2020.7.1.19

I. Introduction

The popularity of the Internet has led to the presence of Web 2.0 technologies, which has given us access to all the topics views of Web 2.0, social media, and any electronic platform in the world over the Internet. Moreover, the spread, adoption, and reliability of social media have created sample opportunities and challenges for the researcher's data. Globally, more than one billion people use social media platforms that generate data on a daily basis and form such data at short intervals. The huge amount of data generated by users periodically is a result of the integration of the backend details and daily activities in this paper. This huge amount of generated data, referred to as "Big Data", will be explained recently.

Data are collected in the form of collections and these Big Data collected can be structured, semi-structured or unstructured in various fields. Such as social networks and health care [1]. Every day, a large number of social media is shown, opening up an area of artificial intelligence and data analysis. It is currently an analysis of social media data using automated learning and data extraction.

It is one of the most important fields. It is an active field of research. It includes the contents of social media such as comments, posts, blogs and references, which contributed to the creation of Big Data on a wide range of different websites or system providers. Basic [2, 3] For example, disclosing information to a particular system can take the views of people and these views lead to improved decision-making processes.

This study deals with the work analyzed in large neural network processes using machine learning methods. In this study, we analyze social media data, including the target ages in the analysis process or the ages that use social media frequently or on a daily basis. Second, we determined the highest percentage of people who use social media, whether they are males or females.

II. Literature Review

Behavioral factors or changes that occur on content before and after publication and how this content is interacted by monitoring and analyzing all individual activities via Facebook, its mechanism uses Difference in Differences (DID)analysis [4].

How to collect information from social media and make a survey to inquire about these data, in return for the result of this examination or query will lead to an expansion in the world of Wikipedia and the most important factor of the data is the chronological order of social relations. Its mechanism is used for temporal statistical analysis [5].

How to deal with the amount of social data available via the Internet. To analyze this data, Hadoop and Spark were used to analyze Big Data, in addition to the proposed algorithm to reduce data-processing costs, its mechanism uses performance analysis [6].

A solution or algorithm for analyzing Big Data using an intra-model reservation algorithm that follows key words for analysis. Its mechanism uses event detection and analysis [7].

The introduction of tips to use social applications that can be used, such as Wikipedia, YouTube, Facebook, and Second Life, to take advantage of this data and support the decision. [8].

How people communicate with each other and must provide platforms for communication among themselves and must encourage customers to participate in these platforms [9].

Six types of flow patterns were used for the video to two-dimensional images. This feature is used for image processing [13].

The results corroborate that the performance of the support vector machine and fuzzy logic algorithms with fewer features offer better accuracy than other machine learning algorithms. These are also computationally less intensive than the other two current techniques [13].

A. *Artificial Neural Networks*

An Artificial Neural Network (ANN) is a computational display that endeavors to represent the parallel idea of the human mind. An (ANN) is a system of very interconnecting handling components (neurons) working in parallel. These components are motivated by organic sensory systems. As in nature, the associations between components to a great extent decide the system work.

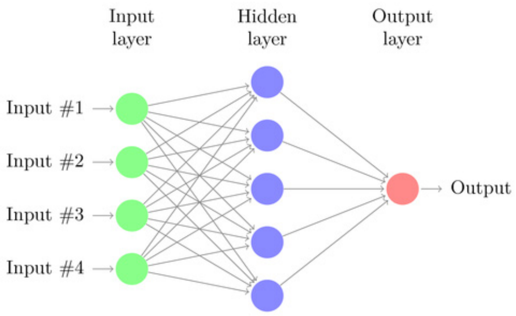


Figure 1: A Typical Neural Network.

This algorithm relies on image processing and optimization of filters to maintain image processing. Using a low-level structure. [14].

Social relationships depend on the different users of social networks to find a link between network theory and communication [15].

The concept of Deep Neural Networks (DNNs) of the most important modern science that focuses on identifying functions of identifying patterns such as images or solving algorithms [10].

Artificial neural networks are inspired by brain modeling called neurons or perceptrons, which are effluents by their internal state activation received from input [11].

Financial transactions are presented and social media data are collected to calculate the credit score. artificial neural networks are used to obtain a general understanding of them [12].

III. Research questions, research objectives, and methodology

A. *Research Questions*

This article attempts to identify a set of questions that must be answered explicitly and clearly. These questions are answered based on Big Data and are analyzed using several methods. Finally, these questions are answered.

Q1: What are the ages that use social media?

In this question, we will look at the ages that are most commonly used for social media.

Q2: Who are the most commonly used people for social media, male or female?

In this question, we will look at who are the most commonly used social media, male or female.

Q3: What are people's interests in writing?

In this question, we will look at the interest of people and what their hobby is in writing.

B. *Research Objectives*

A neural network is a system of devices or software that is designed similar to that of neurons in the human brain. Neural networks are also called artificial neural networks (ANNs). Moreover, artificial intelligence (AI) has a section including deep learning. In this research paper, project data were used, called the social network, and the analysis process was performed using neural network devices. The analysis contains the following data such as number, name, age, source, followers, interests, and Twitter location.

C. Research Methodology

The proposed research methodology used in this study as follows:

Step 1: Select new development project applications. The project of ISBSG. More specifically, the way the Excel.

Step 2: Select the attributes.

Step 3: Filter the data through Excel.

Step 4: Divide the data into two sections, Facebook and Twitter and each of them will choose specific attributes to answer the three questions in this search.

Step 5: Analyze the data based on data clustering.

Step6: Analyze data based on the neural network Model.

Step7: View the results of the research and discussions.

Figure 2 illustrates the detailed research methodology done to achieve the research objectives as follows:

To read the data from the excel, select the required features, with filter projects, can be divided into development or promotion projects, and then detect extreme programs based on selected features will be used techniques, in the end, get results, for example, data analysis based on data clustering and neural networks.

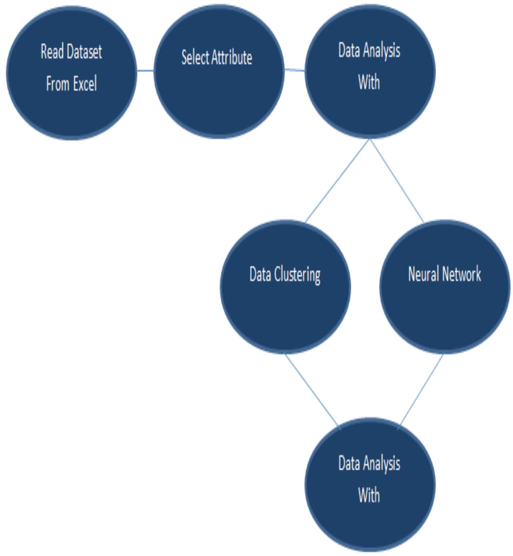


Figure 2: Detailed methodology

IV. Experimental Results

A. Independent Variable

The following Table1 shows the variables used in the analysis of the data will be clarified. These variables depend mainly on the analysis process and are of great importance. The following table shows the variables used:

Environmental data		
S. No	Predictor Variable	Measurement
1	User ID	number
2	Name	liter
3	Age	number
4	Source	liter
5	Followers	number
6	Friends	number
7	Number tweet	number
8	Reply count	number

9	Tweet location	liter
10	Reply location	liter
11	Re-tweet count	number
12	Re-reply count	number
13	Tweet messages	liter

Table 1: Predictor variable of datasets used in the study

The following Table2 explains all the variables used in the data analysis will be explained in detail. These variables are mainly based on the analysis process and are of great importance.

Environmental data		
S. No	Predictor Variable	Description of Data
1	User ID	User ID of each person using social media
2	Name	Name of each person using social media
3	Age	The age of everyone who uses social media
4	Source	Type of social media Facebook and Twitter
5	Followers	Number of followers who use the Twitter platform for a particular person
6	Friends	Number of people using the Facebook platform for a particular person
7	Number tweet	The number of tweets that are used for a particular person
8	Reply count	The number of posts for a particular person
9	Tweet location	From where Tweet has been
10	Reply location	Wherever the comment was made
11	Re-tweet count	How many times have been replying to tweets
12	Re-reply count	How often comments have been answered
13	Tweet messages	Content or Twitter messages

Table 2: Description of data in each predictor variable

B. *Techniques*

The Big Data analytics techniques in the social media context are related to neural networks and data clustering. This section presents important techniques for Big Data analytics with respect to social media data analysis.

C. *Analysis of results using Neural Network*

The data must be analyzed to determine who is using Facebook and Twitter and what are their writing Interests, which are the target ages in this study, we must mention the name of the set of data used is a social network, and the analysis process using the multilayered algorithm.

4.C.1) Analyze data people who use Facebook

In this section, we need to analyze the data of the people who use Facebook and their written interests and the target ages, and the analysis process using a multi-layer algorithm.

Case Processing		Number of Test	Percent
Sample	Training	1338	70.7%
	Testing	554	29.3%
Valid		1892	100.0%
Excluded		0	
Total		1892	

Table 3: Case processing summary

In the previous Table 3, gives us a complete summary of the samples and cases taken from the social data network (Facebook) conducted on the total operations: training, testing, and results can be tracked in percentages.

Information Processing			Results
Input Layer	Factors	1	Age
		2	Message
	Covariates	1	Friends
		2	Rereplycount
	Number of Units		33
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		8
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	Gender
	Number of Units		2
	Activation Function		Softmax
	Error Function		Cross-entropy

Table 4: Network information

In the previous table4, gives us a complete summary of the network information (Facebook). In this table, we summarize the process used, namely the inputs and the process of processing and output. In the first stage. of the input you use the attributes of age, messages, friends, etc. In the second stage, the phase splits data into layers and is divided into 8 layers. In the last stage, this stage is called the output and this stage depends on the main factor in the classification is gender (male or female).

Information Processing		Results
Training	Cross-Entropy Error	823.569
	Percent Incorrect Predictions	30.5%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error
	Training Time	0:00:00.74
Testing	Cross-Entropy Error	323.332
	Percent Incorrect Predictions	26.7%

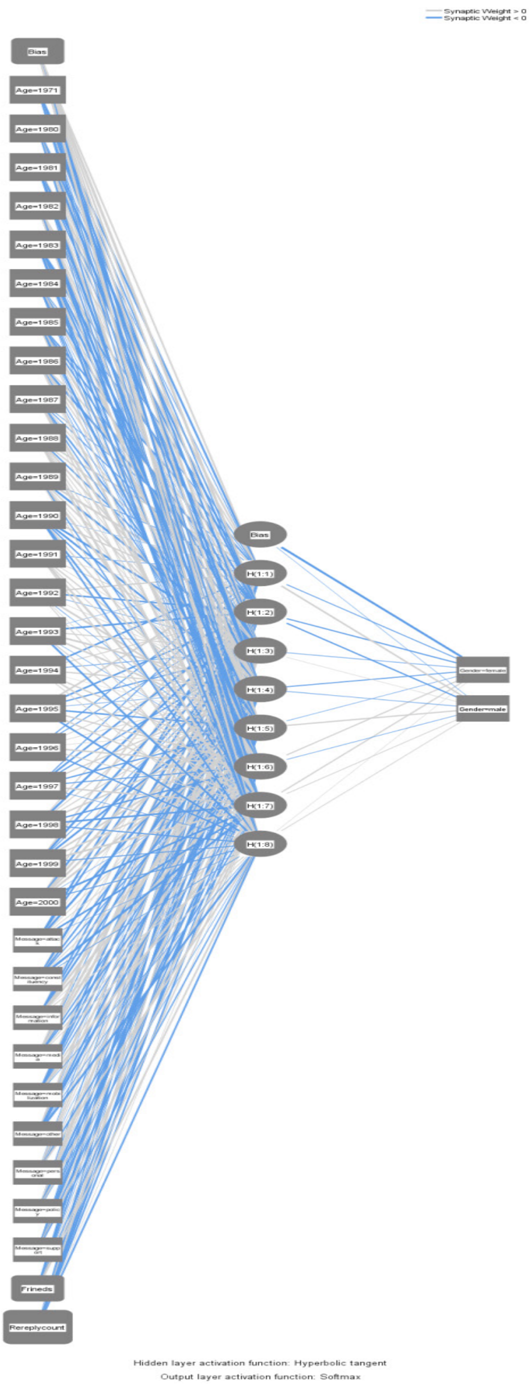
Table 5: Model summary

In the previous table5, gives us a complete summary of the network information (Facebook). This stage is divided into two parts is the first training in this section has several operations, including the percentage of incorrect predictions and time factor in the stage of division, etc. In the second section are the percentage of error tests and the percentage prediction, dependent variables gender.

In the Figure 5, gives us a complete summary of network information (Facebook). In this table, we summarize the process used, namely input, process and outputs. However, at this stage, the text has been converted to visualization for easy data understanding and is always the best way to represent data.in the table below, these images will be summarized to be clarified by typing all data in a table.

Figure 5: Result of data

doi: 10.18638/dialogo.2020.7.1.19



Information Processing	Results
	[Age=1971]
	[Age=1980]
	[Age=1981]
	[Age=1982]
	[Age=1983]
	[Age=1984]
	[Age=1985]
	[Age=1986]
	[Age=1987]
	[Age=1988]
	[Age=1989]
	[Age=1990]
	[Age=1991]
	[Age=1992]
Predictor Input Layer	[Age=1993]
	[Age=1994]
	[Age=1995]
	[Age=1971]
	[Age=1980]
	[Age=1981]
	[Age=1982]
	[Age=1996]
	[Age=1997]
	[Age=1998]
	[Age=1999]
Predictor Input Layer	[Age=2000]
	[Message=attack]
	[Message=constituency]
	[Message=information]
	[Message=media]
	[Message=mobilization]
	[Message=other]
	[Message=personal]
	[Message=policy]
	[Message=support]

Table 6: Parameter estimates

In the previous Table 6, this data was derived from Figure 5. All ages using Facebook were derived from the dataset taken and what interests people in writing. 8 hidden layers were used to linkages and interests, and whether the person using Facebook is male or female.

Sample	Observed	Predicted		
		female	male	Percent Correct
Training	female	5	405	1.2%
	male	3	925	99.7%
	Overall Percent	0.6%	99.4%	69.5%
Testing	female	2	145	1.4%
	male	3	404	99.3%
	Overall Percent	0.9%	99.1%	73.3%

Table 7: Classification

In the previous table 7, two methods of classification were used: training and testing. In both cases, male and female samples were taken, and according to these samples, samples were given in percentages. This table is based on the dependent variable gender.

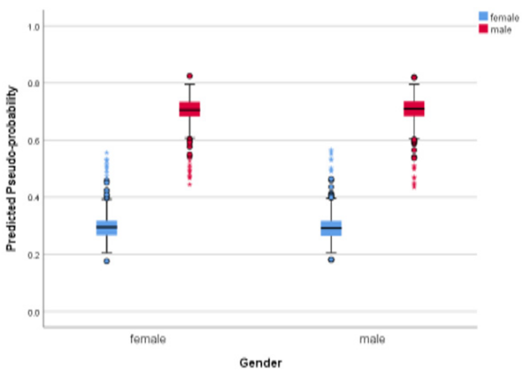


Figure 4: Predicted vs. gender

In the previous figure 4, we took the x and y axis. The x axis is defined as the variable that is called the y-axis and y-axis of the source variable (Facebook). Depending on these two variables, the most commonly used category in Facebook was identified as male or female. The answer is male based on the existing data, and the color (red) indicates the male by drawing.

Predictor	Importance	Normalized Importance
Age	.241	59.7%
Message	.156	38.7%
Friends	.200	49.7%
Re -reply count	.403	100.0%

Table 8: Independent variable importance

In the previous table 8, is one of the most important tables because it shows the most important variables and what is the most important factor. Depending on the response table the response re-reply count is the most important factor among all the factors.

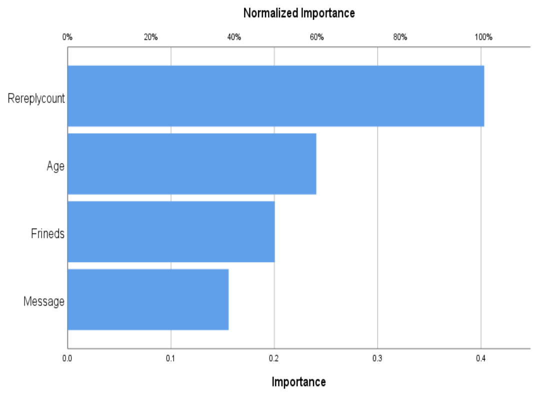


Figure 5: Independent variable of the model summary

In the previous Figure 5, is one of the most important charts because it explains the most important variables and what is the most important factor? Depending on the diagram, the information or factors are explained more clearly than the tables. The re-reply count is the most important factor among all factors.

doi: 10.18638/dialogo.2020.7.1.19

DIALOGO 7:1 (2020)
Multidisciplinary JOURNAL
on the Dialogue between Science and Theology

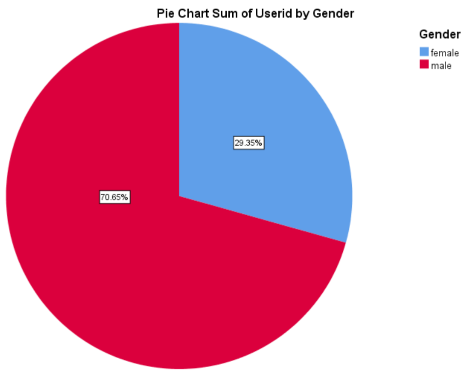


Figure 6: Type of gender

In the previous figure 6, the distribution of gender (male, female) is indicated by percentages. The percentages of males is 70.65% and females are 29.35%.

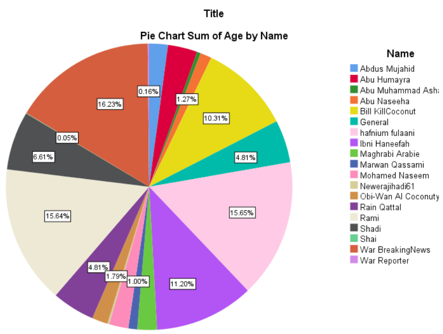


Figure 7: Distribution of age and name

In the previous Figure 7, the distribution of age and name is determined by percentages. The data are distributed depending on the largest category of name, age and given percentages.

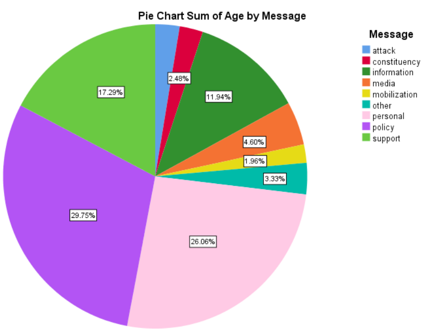


Figure 8: Distribution of message and age

In the previous Figure 8, message distribution and age are determined by percentages. Data is distributed based on the largest category of messages (attack, media, politics, personal, etc.) and age of people using Facebook.

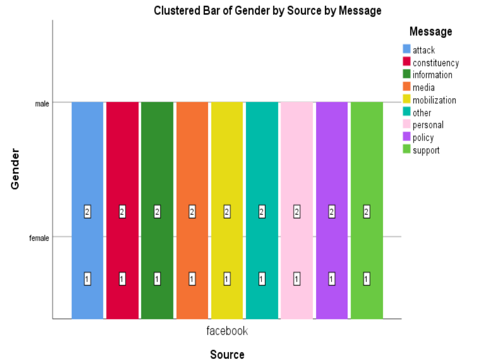


Figure 9: Distribution of message and source

In the previous Figure 9, this picture uses a number of factors have been adopted to find an illustration of the people who use the social networking platform (Facebook), and through the picture shows that the number of males is more used than the number of females, and also explains the interests of people in expressing their opinion through Facebook.

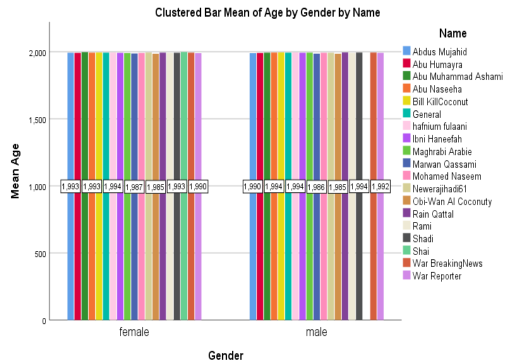


Figure 10: Distribution of gender and name by name

In the previous Figure 10, the picture shows in detail the ages that use Facebook and they are divided into two parts: first, males, and what are the most used ages for Facebook, and the second most used females for Facebook. Other factors, such as name and interests, were also used and the proportions were distributed as shown.

4.C.2) Analyze data on people who use Facebook

k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster. This results in partitioning of the data space into cells.

In this section, the data must be analyzed using another method, namely, the analysis of the K-Means group:

Predictor	Number of Cluster				
	1	2	3	4	5
User id	1297069	2193473	1893767	1594793	1000768
Age	1982	1990	1986	1999	1995
Reply count	0	0	2	0	0
Friends	0	0	2	0	1
Reply count	2	2	2	1	1

Table 9: Initial cluster centers

In the previous Table 9, In this table the sample is taken from each factor more accurately (taking values for each factor). This table is considered the first in terms of taking values was done to five clustering information and created the so-called prototype. One of these factors is age, reply count, friends, etc.

Predictor	Number of Cluster				
	1	2	3	4	5
User id	1290069	2116645	1894248	1590996	1070991
Age	1993	1992	1993	1993	1993
Reply count	4	8	4	4	8
Friends	10	18	11	14	35
Reply count	2	2	2	2	1

Table 10: Final Cluster Centers

Number of cluster		Results
Cluster	1	446.000
	2	246.000
	3	493.000
	4	463.000
	5	244.000
Valid		1892.000
Missing		.000

Table 11: Number of cases in each cluster

4.C.3) Analyze data on people who use Twitter

In this section, we need to analyze the data of the people who use Twitter and their written interests and the target ages, and the analysis process using a multi-layer algorithm:

Case Processing		Number of Test	Percent
Sample	Training	929	70.5%
	Testing	389	29.5%
Valid		1318	100.0%
Excluded		0	
Total		1318	

Table 12: Case processing summary

In the previous Table12, a complete summary of the samples and case taken from the social data network (Twitter) conducted on the total operations: training, testing and results can be tracked in percentages.

	Information Processing		Results
Input Layer	Factors	1	Age
		2	Message
	Covariates	1	Followers
		2	Retweetcount
	Number of Units		31
	Rescaling Method for Covariates		Standardized

doi: 10.18638/dialogo.2020.7.1.19

DIALOGO 7:1 (2020)
Multidisciplinary JOURNAL
on the Dialogue between Science and Theology

Hidden layer(s)	Number of Hidden Layers	1
	Number of Units in Hidden Layer 1[a]	8
	Activation Function	Hyperbolic tangent
Output Layer	Dependent Variables 1	Gender
	Number of Units	2
	Activation Function	Softmax
	Error Function	Cross-entropy

Table 13: Network information

In the previous Table13, a complete summary of the network information (Twitter). In this table, we summarize the process used, namely the inputs and the process of processing and output. In the first stage. of the input, you use the attributes of age, messages, followers, etc. In the second stage, this phase splits data into layers and is divided into 8 layers. In the last stage, this stage is called the output and this stage depends on the main factor in the classification is gender (male or female).

	Information Processing	Results
Training	Cross Entropy Error	534.896
	Percent Incorrect Predictions	27.3%
	Stopping Rule Used	1 consecutive step(s) with no decrease in error
	Training Time	0:00:00.92
Testing	Cross Entropy Error	230.821
	Percent Incorrect Predictions	28.3%

Table 14: Model summary

In the previous table14 gives us a complete summary of the network information (twitter). In this stage is divided into two parts is the first training in this section has several operations, including the percentage of incorrect predictions and time factor in the stage of division, etc. In the second section are the percentage of error tests and the percentage prediction, dependent variable is gender. dependent variable is Gender.

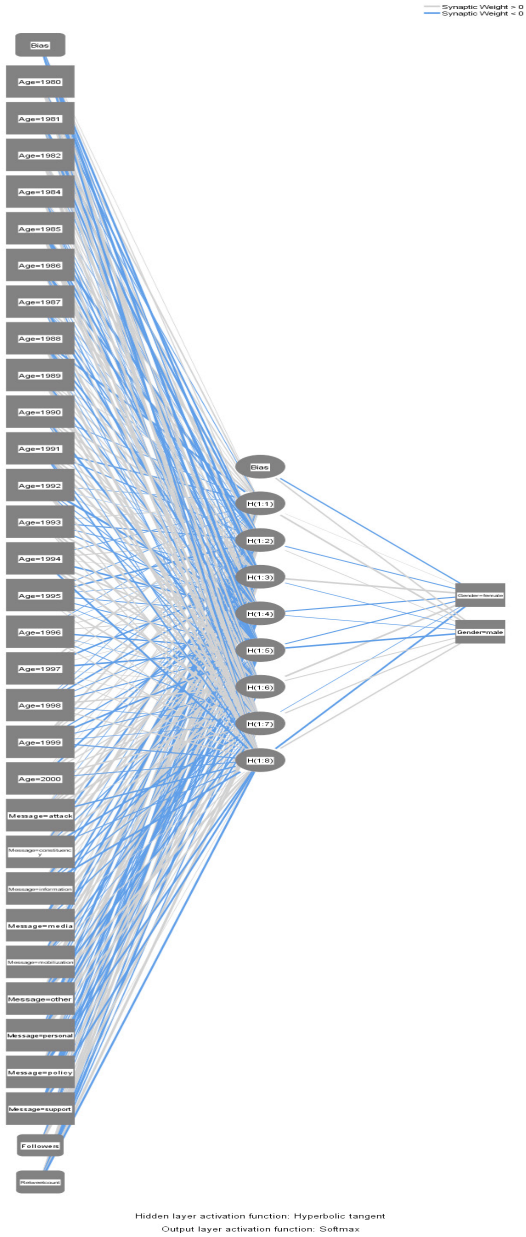


Figure 11: Result of data

In the previous Figure 11, a complete summary of network information (Twitter) is given. In this figure we summarize the process used, namely input, process and outputs. However, at this stage, the text has been converted to visualization for easy data understanding and is always the best way to represent data. In the table below, these images summarized to be clarified by typing all data in a table.

Information Processing	Results
Predictor Input Layer	[Age=1980]
	[Age=1981]
	[Age=1982]
	[Age=1984]
	[Age=1985]
	[Age=1987]
	[Age=1988]
	[Age=1989]
	[Age=1990]
	[Age=1991]
	[Age=1992]
	[Age=1993]
	[Age=1994]
	[Age=1995]
	[Age=1996]
	[Age=1997]
	[Age=1998]
	[Age=1999]
	[Age=2000]
	[Message=attack]
	[Message=constituency]
	[Message=information]
	[Message=media]
	[Message=mobilization]
	[Message=other]
	[Message=personal]
	[Message=policy]
	[Message=support]
	Followers
Predictor Input Layer	Retweet count

Table 15: Parameter estimates

In the previous Table 15, this data was derived from Figure 11. All ages using Twitter were derived from the dataset taken and what interests people in writing. 8 hidden layers were used to link ages and interests, and whether the person using Twitter was male or female.

Sample	Observed	Predicted		
		female	male	Percent Correct
Training	female	7	244	2.8%
	male	10	668	98.5%
	Overall Percent	1.8%	98.2%	72.7%
Testing	female	2	108	1.8%
	male	2	277	99.3%
	Overall Percent	1.0%	99.0%	71.7%

Table 16: Classification

In the previous table 16 this data was derived from the figure 11. All ages using twitter were derived from the dataset taken and what interests people in writing. 8 hidden layers were used to link ages and interests, and whether the person using twitter was male or female.

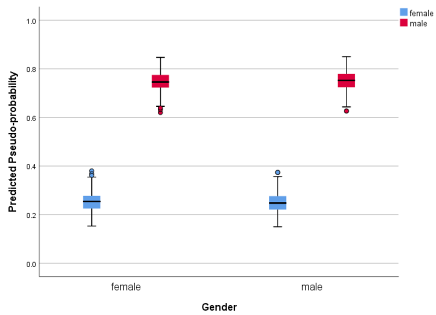


Figure 12: Predicted vs. gender

In the previous figure 12, we took the x and y axis. The x axis is defined as the variable that is called the y-axis and y-axis of the source variable (twitter). Depending on these two variables, the most commonly used category in Twitter was identified as male or female. The answer is male based on the existing data, and the color (red) indicates the male by drawing.

doi: 10.18638/dialogo.2020.7.1.19

Predictor	Importance	Normalized Importance
Age	.347	99.2%
Message	.349	100.0%
Followers	.144	41.1%
Retweetcount	.161	46.0%

Table 17: Independent variable importance

In the previous Table 17, is one of the most important tables because it shows the most important variables and what is the most important factor. Depending on the table, the message is the most important factor among all the factors.

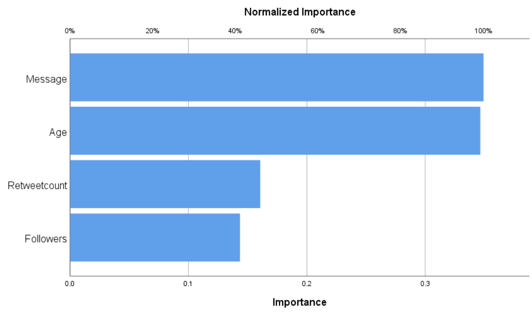


Figure 13: Independent variable of model summary

In the previous Figure 13, is one of the most important charts because it explains the most important variables and what is the most important factor? Depending on the diagram, the information or factors are explained more clearly than the tables. The message is the most important factor among all factors.

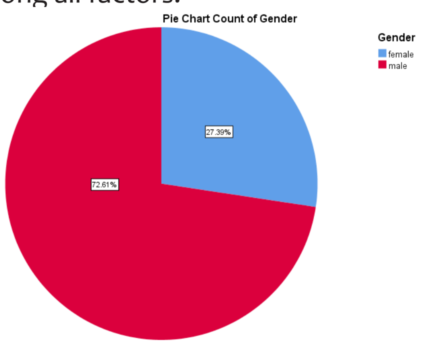


Figure 14: Type of gender

In the previous Figure 14, the distribution of gender (male, female) is indicated by percentages. The percentages of males is 72.61 and females are 27.39.

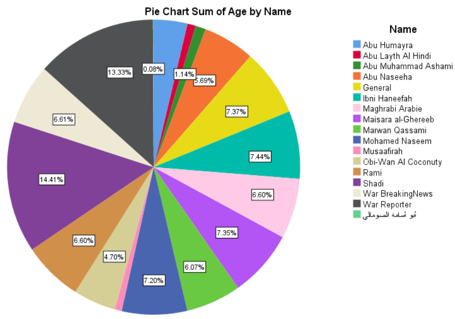


Figure 15: Distribution of age and name

In the previous Figure 15, the distribution of age and name is determined by percentages. The data are distributed depending on the largest category of name, age and given percentages.

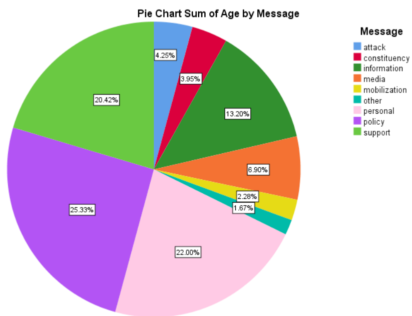


Figure 16: Distribution of message and age

In the previous Figure 16, message distribution and age are determined by percentages. data are distributed based on the largest category of messages (attack, media, politics, personal, etc.) and age of people using Twitter.

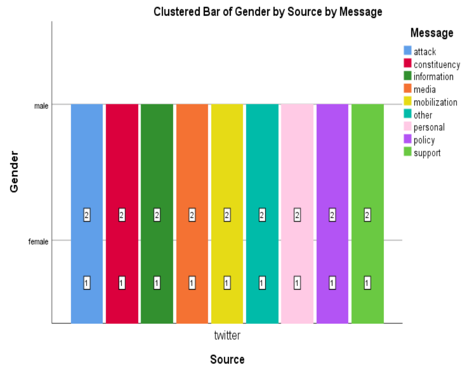


Figure 17: Distribution of message and source

In the previous Figure 17, In this picture, a number of factors have been adopted to find an illustration of the people who use the social networking platform (Twitter), and through the picture shows that the number of males is more used than the number of females, and also explains the interests of people in expressing their opinion through Twitter.

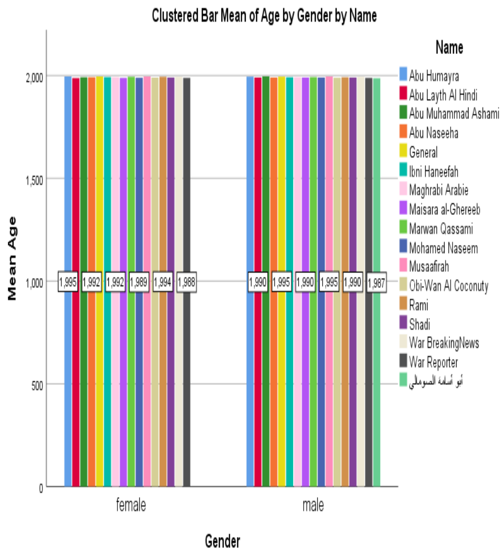


Figure 18: Distribution of gender and name

In the previous Figure 19, the picture shows in detail the ages that use Twitter and they are divided into two parts: first, males, and what are the most used ages for Twitter, and the second most used females for Twitter. Other factors, such as name and interests, were also used and the proportions were distributed as shown.

4.C.4) Analyze data on people who use Twitter

In this section, the data must be analyzed using another method, namely, the analysis of the K-Means group:

Predictor	Number of Cluster				
	1	2	3	4	5
Userid	1000455	1892673	2193232	1596443	1297138
Age	1985	1997	1998	1988	1998
Number tweet	607	743	16688	16688	656
Followers	158	392	29209	29209	139
Retweet count	0	0	0	0	0

Table 18: Initial cluster centers

In the previous Table 18, In this table the sample is taken from each factor more accurately (taking values for each factor). This table is considered the first in terms of taking values. was done to five clustering information and created the so-called prototype. Other factors are age, number of tweets, followers, etc.

Predictor	Number of Cluster				
	1	2	3	4	5
Userid	1085193	1897855	2117932	1590905	1290858
Age	1991	1992	1991	1991	1991
Number tweet	3390	3267	3607	3172	3175
Followers	3338	3142	3365	2948	2786
Retweet count	0	0	0	0	0

Table 19: Final cluster centers

Number of cluster		Results
Cluster	1	162.000
	2	338.000
	3	144.000
	4	335.000
	5	339.000
Valid		1318.000
Missing		.000

Table 20: Number of cases in each cluster

Conclusion

The enormous information about life on the Internet has evolved along with the advancement of computer tools as a way of critical experiences in human behavior. It has been constantly raised by companies, people, and governments. The paper discussed several of their fundamentals by analyzing Big Data. This paper was divided into two sections in terms of social media to Facebook and Twitter and it was found that the age group based on the data used ranges from the twenties to the thirties. Second, the people were divided into two parts in terms of communication, mail to Facebook and Twitter have been reaching the people most frequently used for social networking based on the data used are males also were divided. Third of people into two parts in terms of social media to Facebook and Twitter were to reach most of the concerns of people confined attack, personal, support, constituency, policy, media, information, mobilization, and others. Finally, we see the difficulties of open research in the testing of Big Data because it depends on the type of data and how to analyze it.

References

[1] Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Khan, S. U. The rise of "Big Data" on cloud computing: Review and open research issues. Information systems, 2015, 47, 98-115.

[2] Kwon, O., Lee, N., & Shin, B. Data quality management, data usage experience and acquisition intention of Big Data analytics. *International Journal of Information Management*, 2014; 34(3), 387-394.

[3] Lyu, K., & Kim, H. Sentiment analysis using word polarity of social media. *Wireless Personal Communications*, 2016; 89(3), 941-958.

[4] Grinberg, N., Dow, P. A., Adamic, L. A., Naaman, M.. "Changes in engagement before and after posting to facebook". Paper presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems.

[5] Do, N., Rahayu, W., Torabi, T. A query expansion approach for social media data extraction. *International Journal of Web and Grid Services*, 2016; 12(4), 418-441.

[6] Santos, M. C., Meira, W., Guedes, D., Almeida, V. F. (2016). "Faster: a low overhead framework for massive data analysis." Paper presented at the Cluster, Cloud and Grid Computing (CCGrid), 2016; 16th IEEE/ACM International Symposium on.

[7] Li, J., Rao, Y., Jin, F., Chen, H., Xiang, X. "Multi-label maximum entropy model for social emotion classification over short text." *Neurocomputing*, 2016; 210, 247-256.

[8] Kaplan, Andreas M., Michael Haenlein. "Users of the world, unite! The challenges and opportunities of Social Media". *Business horizons*, 2010; 53 (1), 59-68.

[9] Mangold, W. Glynn, and David J. Faulds. "Social media: The new hybrid element of the promotion mix". *Business horizons,* 2009; 52 (4): 357-365.

[10] Nguyen, Anh, Jason Yosinski, and Jeff Clune. "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 427-436. 2015.

[11] Fulcher, J. "Computational intelligence: an introduction Computational intelligence: a compendium" (pp. 3-78): *Springer*, 2008.

[12] Ghiassi, M., Zimbra, D., Lee, S. "Targeted twitter sentiment analysis for brands

using supervised feature engineering and the dynamic architecture for artificial neural networks." *Journal of Management Information Systems*, 2016; 33(4), 1034-1058.

[13] Shanthi, C., Pappa, N. "An artificial intelligence based improved classification of two-phase flow patterns with feature extracted from acquired images." *ISA transactions,* 2017; 68, 425-432.

[14] Fan, Z., Bi, D., He, L., Shiping, M., Gao, S., Li, C. "Low-level structure feature extraction for image processing via stacked sparse denoising autoencoder." *Neurocomputing*, 2017; 243, 12-20.

[15] Serrat, O. "Social network analysis Knowledge solutions" (pp. 39-43): *Springer*, 2017.

Biography

Dr. Ahed Alkhatib has finished his PhD from Cambell State University in 2011. I am currently working as a clinical researcher at faculty of medicine, Jordan University of Science and technology. Over the time, I have published more than 200 articles in various medical fields including neurosciences, pharmacology, and diabetes. My approaches in research include the involvement of philosophy of science in research which gives looking, and thinking in depth. I have developed several hypotheses in medicine such as the role of white matter in initiating diseases such as diabetes. In microbiology, I have demonstrated that prokaryotic cells and eukaryotic cells are similar in producing cell cycle proteins which can participate in autoimmunity diseases.

Research interest: Setting medical hypotheses and writing book in different fields, of which two books have already been written and distributed in the world market.

Other interests include:

- Neurology
- Pharmacology
- Philosophy of science
- Tumor research
- Behavioral biology
- Emergency medicine